

SSB ban - NHANES

Sanjay Basu

5/28/2019

NHANES import

extract and define needed variables from NHANES 2011-2016 note the CKD-EPI equation is used to define CKD, http://nephron.com/epi_equation ; our justification for this decision was the results from: <https://jamanetwork.com/journals/jama/fullarticle/1151529>

```
demo_g = nhanes('DEMO_G') %>%
  mutate(age = RIDAGEYR,
         female = (RIAGENDR==2),
         white = (RIDRETH3==3),
         black = (RIDRETH3==4),
         asian = (RIDRETH3==6),
         other = ((RIDRETH3!=3)&(RIDRETH3!=4)&(RIDRETH3!=6)),
         hisp = (RIDRETH3<=2)) %>%
  select(SEQN, age, female, white, black, asian, other, hisp)
```

```
## Processing SAS dataset DEMO_G      ..
```

```
hiq_g = nhanes('HIQ_G') %>%
  mutate(privins = (HIQ031A==14),
         privins = replace_na(privins, 0)) %>%
  select(SEQN, privins)
```

```
## Processing SAS dataset HIQ_G      ..
```

```
ocq_g = nhanes('OCQ_G') %>%
  filter((OCD231 <21)&(OCD150 ==1)) %>%
  mutate(mgmt = (OCD241<=7),
         blcol = (((OCD241>=12)&(OCD241<=15))|((OCD241>=18)&(OCD241<=22))),
         whcol = (((OCD241>=8)&(OCD241<=11))|((OCD241>=16)&(OCD241<=17)))) %>%
  select(SEQN, mgmt, blcol, whcol)
```

```
## Processing SAS dataset OCQ_G      ..
```

```
bmx_g = nhanes('BMX_G') %>%
  mutate(bmi = BMXBMI,
         wtkg = BMXWT,
         htm = BMXHT/100) %>%
  select(SEQN, bmi, wtkg, htm)
```

```
## Processing SAS dataset BMX_G      ..
```

```
mcq_g = nhanes('MCQ_G') %>%
  mutate(chd = (MCQ160B==1) | (MCQ160C==1) | (MCQ160D==1) | (MCQ160E==1),
         str = (MCQ160F==1)) %>%
  select(SEQN, chd, str)
```

```
## Processing SAS dataset MCQ_G      ..
```

```
rxq_drug = nhanes('RXQ_DRUG')
```

```
## Processing SAS dataset RXQ_DRUG  ..
```

```
rxq_rx_g = nhanes('RXQ_RX_G') %>%
  left_join(rxq_drug, by = "RXDDRGID") %>%
  filter(RXDDCN1B=="ANTIDIABETIC AGENTS") %>%
  group_by(SEQN) %>%
  mutate(dmmed = 1) %>%
  select(SEQN, dmmed)
```

```
## Processing SAS dataset RXQ_RX_G  ..
```

```
## Warning: Column `RXDDRGID` joining factor and character vector, coercing
## into character vector
```

```
## Warning: Column `RXDDRGID` has different attributes on LHS and RHS of join
```

```
ghb_g = nhanes('GHB_G') %>%
  select(SEQN, LBXGH)
```

```
## Processing SAS dataset GHB_G      ..
```

```
glu_g = nhanes('GLU_G') %>%
  select(SEQN, LBXGLU)
```

```
## Processing SAS dataset GLU_G      ..
```

```
ogtt_g = nhanes('OGTT_G') %>%
  select(SEQN, LBXGLT)
```

```
## Processing SAS dataset OGTT_G     ..
```

```
diq_g = nhanes('DIQ_G') %>%
  full_join(ghb_g, by="SEQN") %>%
  full_join(glu_g, by="SEQN") %>%
  full_join(ogtt_g, by="SEQN") %>%
  full_join(rxq_rx_g, by="SEQN") %>%
  mutate(dm = ((LBXGH>=6.5) | (LBXGLU>=126) | (LBXGLT>=200) | (DIQ010==1) | (dmmed==1)),
         dm = replace_na(dm, 0)) %>%
  select(SEQN, dm)
```

```
## Processing SAS dataset DIQ_G      ..
```

```
biopro_g = nhanes('BIOPRO_G') %>%
  select(SEQN, LBXSCR)
```

Processing SAS dataset BIOPRO_G ..

```
kiq_u_g = nhanes('KIQ_U_G') %>%
  full_join(biopro_g, by="SEQN") %>%
  full_join(demo_g, by="SEQN") %>%
  mutate(scr = LBXSCR,
         kappa = 0.7*female + 0.9 *(1-female),
         alpha = -0.329*female -0.411*(1-female),
         sk = scr/kappa,
         minsk = sk*(sk<1)+1*(sk>=1),
         maxsk = sk*(sk>=1)+1*(sk<1),
         egfr = 141 *(minsk^alpha) *
           (maxsk^(-1.209))*
           (0.993^age)*
           (1.018*female+1*(1-female))*
           (1.159*black+1*(1-black)),
         ckd = (egfr<90)) %>%
  select(SEQN, ckd)
```

Processing SAS dataset KIQ_U_G ..

```
ohxref_g = nhanes('OHXREF_G') %>%
  mutate(dental = (OHAROCDT==1)&(OHAROCGP==1)) %>%
  select(SEQN, dental)
```

Processing SAS dataset OHXREF_G ..

```
demo_h = nhanes('DEMO_H') %>%
  mutate(age = RIDAGEYR,
         female = (RIAGENDR==2),
         white = (RIDRETH3==3),
         black = (RIDRETH3==4),
         asian = (RIDRETH3==6),
         other = ((RIDRETH3!=3)&(RIDRETH3!=4)&(RIDRETH3!=6)),
         hisp = (RIDRETH3<=2)) %>%
  select(SEQN, age, female, white, black, asian, other, hisp)
```

Processing SAS dataset DEMO_H ..

```
hiq_h = nhanes('HIQ_H') %>%
  mutate(privins = (HIQ031A==14),
         privins = replace_na(privins, 0)) %>%
  select(SEQN, privins)
```

Processing SAS dataset HIQ_H ..

```
ocq_h = nhanes('OCQ_H') %>%
  filter((OCQ260 ==1)&(OCD150 ==1)) %>%
  mutate(mgmt = NA,
         blcol = NA,
         whcol = NA) %>%
  select(SEQN, mgmt, blcol, whcol)
```

Processing SAS dataset OCQ_H ..

```
bmx_h = nhanes('BMX_H') %>%
  mutate(bmi = BMXBMI,
         wtkg = BMXWT,
         htm = BMXHT/100) %>%
  select(SEQN, bmi, wtkg, htm)
```

Processing SAS dataset BMX_H ..

```
mcq_h = nhanes('MCQ_H') %>%
  mutate(chd = (MCQ160B==1) | (MCQ160C==1) | (MCQ160D==1) | (MCQ160E==1),
         str = (MCQ160F==1)) %>%
  select(SEQN, chd, str)
```

Processing SAS dataset MCQ_H ..

```
rxq_drug = nhanes('RXQ_DRUG')
```

Processing SAS dataset RXQ_DRUG ..

```
rxq_rx_h = nhanes('RXQ_RX_H') %>%
  left_join(rxq_drug, by = "RXDDRGID") %>%
  filter(RXDDCN1B=="ANTIDIABETIC AGENTS") %>%
  group_by(SEQN) %>%
  mutate(dmmed = 1) %>%
  select(SEQN, dmmed)
```

Processing SAS dataset RXQ_RX_H ..

Warning: Column `RXDDRGID` joining factor and character vector, coercing
into character vector

Warning: Column `RXDDRGID` has different attributes on LHS and RHS of join

```
ghb_h = nhanes('GHB_H') %>%
  select(SEQN, LBXGH)
```

Processing SAS dataset GHB_H ..

```
glu_h = nhanes('GLU_H') %>%
  select(SEQN, LBXGLU)
```

```
## Processing SAS dataset GLU_H    ..
```

```
ogtt_h = nhanes('OGTT_H') %>%
  select(SEQN, LBXGLT)
```

```
## Processing SAS dataset OGTT_H   ..
```

```
diq_h = nhanes('DIQ_H') %>%
  full_join(ghb_h, by="SEQN") %>%
  full_join(glu_h, by="SEQN") %>%
  full_join(ogtt_h, by="SEQN") %>%
  full_join(rxq_rx_h, by="SEQN") %>%
  mutate(dm = ((LBXGH>=6.5) | (LBXGLU>=126) | (LBXGLT>=200) | (DIQ010==1) | (dmmed==1)),
         dm = replace_na(dm, 0)) %>%
  select(SEQN, dm)
```

```
## Processing SAS dataset DIQ_H    ..
```

```
biopro_h = nhanes('BIOPRO_H') %>%
  select(SEQN, LBXSCR)
```

```
## Processing SAS dataset BIOPRO_H ..
```

```
kiq_u_h = nhanes('KIQ_U_H') %>%
  full_join(biopro_h, by="SEQN") %>%
  full_join(demo_h, by="SEQN") %>%
  mutate(scr = LBXSCR,
         kappa = 0.7*female + 0.9*(1-female),
         alpha = -0.329*female -0.411*(1-female),
         sk = scr/kappa,
         minsk = sk*(sk<1)+1*(sk>=1),
         maxsk = sk*(sk>=1)+1*(sk<1),
         egfr = 141*(minsk^alpha)*
           (maxsk^(-1.209))*
           (0.993^age)*
           (1.018*female+1*(1-female))*
           (1.159*black+1*(1-black)),
         ckd = (egfr<90)) %>%
  select(SEQN, ckd)
```

```
## Processing SAS dataset KIQ_U_H  ..
```

```
ohxref_h = nhanes('OHXREF_H') %>%
  mutate(dental = (OHAROCDT==1)&(OHAROCGP==1)) %>%
  select(SEQN, dental)
```

```
## Processing SAS dataset OHXREF_H ..
```

```
demo_i = nhanes('DEMO_I') %>%
  mutate(age = RIDAGEYR,
         female = (RIAGENDR==2),
         white = (RIDRETH3==3),
         black = (RIDRETH3==4),
         asian = (RIDRETH3==6),
         other = ((RIDRETH3!=3)&(RIDRETH3!=4)&(RIDRETH3!=6)),
         hisp = (RIDRETH3<=2)) %>%
  select(SEQN, age, female, white, black, asian, other, hisp)
```

```
## Processing SAS dataset DEMO_I    ..
```

```
hiq_i = nhanes('HIQ_I') %>%
  mutate(privins = (HIQ031A==14),
         privins = replace_na(privins, 0)) %>%
  select(SEQN, privins)
```

```
## Processing SAS dataset HIQ_I    ..
```

```
ocq_i = nhanes('OCQ_I') %>%
  filter((OCQ260 ==1)&(OCD150 ==1)) %>%
  mutate(mgmt = NA,
         blcol = NA,
         whcol = NA) %>%
  select(SEQN, mgmt, blcol, whcol)
```

```
## Processing SAS dataset OCQ_I    ..
```

```
bmx_i = nhanes('BMX_I') %>%
  mutate(bmi = BMXBMI,
         wtkg = BMXWT,
         htm = BMXHT/100) %>%
  select(SEQN, bmi, wtkg, htm)
```

```
## Processing SAS dataset BMX_I    ..
```

```
mcq_i = nhanes('MCQ_I') %>%
  mutate(chd = (MCQ160B==1) | (MCQ160C==1) | (MCQ160D==1) | (MCQ160E==1),
         str = (MCQ160F==1)) %>%
  select(SEQN, chd, str)
```

```
## Processing SAS dataset MCQ_I    ..
```

```
rxq_drug = nhanes('RXQ_DRUG')
```

```
## Processing SAS dataset RXQ_DRUG ..
```

```
rxq_rx_i = nhanes('RXQ_RX_I') %>%
  left_join(rxq_drug, by = "RXDDRGID") %>%
  filter(RXDDCN1B=="ANTIDIABETIC AGENTS") %>%
  group_by(SEQN) %>%
  mutate(dmmed = 1) %>%
  select(SEQN,dmmed)
```

```
## Processing SAS dataset RXQ_RX_I    ..
```

```
## Warning: Column `RXDDRGID` joining factor and character vector, coercing
## into character vector
```

```
## Warning: Column `RXDDRGID` has different attributes on LHS and RHS of join
```

```
ghb_i = nhanes('GHB_I') %>%
  select(SEQN, LBXGH)
```

```
## Processing SAS dataset GHB_I      ..
```

```
glu_i = nhanes('GLU_I') %>%
  select(SEQN, LBXGLU)
```

```
## Processing SAS dataset GLU_I      ..
```

```
ogtt_i = nhanes('OGTT_I') %>%
  select(SEQN, LBXGLT)
```

```
## Processing SAS dataset OGTT_I     ..
```

```
diq_i = nhanes('DIQ_I') %>%
  full_join(ghb_i, by="SEQN") %>%
  full_join(glu_i, by="SEQN") %>%
  full_join(ogtt_i, by="SEQN") %>%
  full_join(rxq_rx_i, by="SEQN") %>%
  mutate(dm = ((LBXGH>=6.5) | (LBXGLU>=126) | (LBXGLT>=200) | (DIQ010==1) | (dmmed==1)),
         dm = replace_na(dm, 0)) %>%
  select(SEQN,dm)
```

```
## Processing SAS dataset DIQ_I      ..
```

```
biopro_i = nhanes('BIOPRO_I') %>%
  select(SEQN, LBXSCR)
```

```
## Processing SAS dataset BIOPRO_I   ..
```

```

kiq_u_i = nhanes('KIQ_U_I') %>%
  full_join(biopro_i, by="SEQN") %>%
  full_join(demo_i, by="SEQN") %>%
  mutate(scr = LBXSCR,
         kappa = 0.7*female + 0.9 *(1-female),
         alpha = -0.329*female -0.411*(1-female),
         sk = scr/kappa,
         minsk = sk*(sk<1)+1*(sk>=1),
         maxsk = sk*(sk>=1)+1*(sk<1),
         egfr = 141 *(minsk^alpha) *
           (maxsk^(-1.209))*
           (0.993^age)*
           (1.018*female+1*(1-female))*
           (1.159*black+1*(1-black)),
         ckd = (egfr<90)) %>%
  select(SEQN, ckd)

```

```
## Processing SAS dataset KIQ_U_I    ..
```

```

ohxref_i = nhanes('OHXREF_I') %>%
  mutate(dental = (OHAROCDT==1)&(OHAROCGP==1)) %>%
  select(SEQN, dental)

```

```
## Processing SAS dataset OHXREF_I    ..
```

NHANES join

join and limit to employed pop by occ class, starting with ocq and left joins to limit to those employed in priv sec

```

nhanes11 = ocq_g %>%
  left_join(demo_g, by="SEQN") %>%
  left_join(hiq_g, by="SEQN") %>%
  left_join(bmx_g, by="SEQN") %>%
  left_join(mcq_g, by="SEQN") %>%
  left_join(diq_g, by="SEQN") %>%
  left_join(kiq_u_g, by="SEQN") %>%
  left_join(ohxref_g, by="SEQN")

nhanes13 = ocq_h %>%
  left_join(demo_h, by="SEQN") %>%
  left_join(hiq_h, by="SEQN") %>%
  left_join(bmx_h, by="SEQN") %>%
  left_join(mcq_h, by="SEQN") %>%
  left_join(diq_h, by="SEQN") %>%
  left_join(kiq_u_h, by="SEQN") %>%
  left_join(ohxref_h, by="SEQN")

nhanes15 = ocq_i %>%
  left_join(demo_i, by="SEQN") %>%
  left_join(hiq_i, by="SEQN") %>%

```



```
nhanes = bind_rows(list(nhanes11, nhanes13, nhanes15))
```

```
## Warning in bind_rows(x, .id): Vectorizing 'labelled' elements may not
## preserve their attributes
```

```
summary(nhanes)
```

```
##      SEQN      mgmt      blcol      whcol
## Min.   :62161  Mode :logical  Mode :logical  Mode :logical
## 1st Qu.:69364  FALSE:2290  FALSE:1640    FALSE:1864
## Median :78169  TRUE :607   TRUE :1257    TRUE :1033
## Mean   :77766  NA's :5467  NA's :5467    NA's :5467
## 3rd Qu.:85524
## Max.   :93702
##
##      age      female      white      black
## Min.   :16.00  Mode :logical  Mode :logical  Mode :logical
## 1st Qu.:29.00  FALSE:4476    FALSE:5457    FALSE:6554
## Median :40.00  TRUE :3888    TRUE :2907    TRUE :1810
## Mean   :40.91
## 3rd Qu.:52.00
## Max.   :80.00
##
##      asian      other      hisp      privins
## Mode :logical  Mode :logical  Mode :logical  Min.   :0.0000
## FALSE:7173    FALSE:5908    FALSE:6181    1st Qu.:0.0000
## TRUE :1191     TRUE :2456    TRUE :2183    Median :1.0000
##                                     Mean    :0.6111
##                                     3rd Qu.:1.0000
##                                     Max.    :1.0000
##
##      bmi      wtkg      htm      chd
## Min.   :14.10  Min.   : 32.80  Min.   :1.353  Mode :logical
## 1st Qu.:23.80  1st Qu.: 66.40  1st Qu.:1.612  FALSE:7630
## Median :27.60  Median : 78.60  Median :1.681  TRUE :210
## Mean   :28.81  Mean   : 81.86  Mean   :1.683  NA's :524
## 3rd Qu.:32.30  3rd Qu.: 92.80  3rd Qu.:1.753
## Max.   :80.60  Max.   :216.10  Max.   :2.027
## NA's   :353    NA's   :350    NA's   :351
##
##      str      dm      ckd      dental
## Mode :logical  Min.   :0.0000  Mode :logical  Mode :logical
## FALSE:7768    1st Qu.:0.0000  FALSE:5680    FALSE:44
## TRUE :72      Median :0.0000  TRUE :1949    TRUE :886
## NA's :524     Mean   :0.1451  NA's :735     NA's :7434
##                                     3rd Qu.:0.0000
##                                     Max.    :1.0000
##
```

Imputation

10-fold multiple imputation w/ chained equations

```
set.seed(123)
plan(multiprocess)
start = Sys.time()
imp_nhanes = future_map(rep(1, 10), ~mice(data = nhanes, m = ., maxit = 100, printFlag = FALSE))
```

```
## Warning: Number of logged events: 1000
## Warning: Number of logged events: 1000
## Warning: Number of logged events: 1000
## Warning: Number of logged events: 1000
## Warning: Number of logged events: 1000
## Warning: Number of logged events: 1000
## Warning: Number of logged events: 1000
## Warning: Number of logged events: 1000
## Warning: Number of logged events: 1000
## Warning: Number of logged events: 1000
```

```
end = Sys.time() - start
print(end)
```

```
## Time difference of 6.160395 mins
```

```
imp_nhanes_complete = map(imp_nhanes, mice::complete)
imp_nhanes = map2(.x = seq(1,10), .y = imp_nhanes_complete, ~mutate(.y, imp_id = as.character(.x)))
imp_nhanes = bind_rows(imp_nhanes)
summary(imp_nhanes)
```

```
##      SEQN      mgmt      blcol      whcol
##  Min.   :62161  Min.   :0.0000  Min.   :0.0000  Min.   :0.0000
## 1st Qu.:69364  1st Qu.:0.0000  1st Qu.:0.0000  1st Qu.:0.0000
## Median :78169  Median :0.0000  Median :0.0000  Median :1.0000
## Mean   :77766  Mean   :0.4966  Mean   :0.4907  Mean   :0.5009
## 3rd Qu.:85524  3rd Qu.:1.0000  3rd Qu.:1.0000  3rd Qu.:1.0000
## Max.   :93702  Max.   :1.0000  Max.   :1.0000  Max.   :1.0000
##      age      female      white      black
##  Min.   :16.00  Mode :logical  Mode :logical  Mode :logical
## 1st Qu.:29.00  FALSE:44760   FALSE:54570   FALSE:65540
## Median :40.00  TRUE :38880    TRUE :29070    TRUE :18100
## Mean   :40.91
## 3rd Qu.:52.00
## Max.   :80.00
##      asian      other      hisp      privins
##  Mode :logical  Mode :logical  Mode :logical  Min.   :0.0000
## FALSE:71730   FALSE:59080   FALSE:61810   1st Qu.:0.0000
## TRUE :11910    TRUE :24560    TRUE :21830   Median :1.0000
##                                     Mean   :0.6111
##                                     3rd Qu.:1.0000
##                                     Max.   :1.0000
##      bmi      wtkg      htm      chd
##  Min.   :14.10  Min.   : 32.80  Min.   :1.353  Min.   :0.00000
```

```
## 1st Qu.:23.70 1st Qu.: 65.80 1st Qu.:1.611 1st Qu.:0.00000
## Median :27.40 Median : 77.70 Median :1.682 Median :0.00000
## Mean :28.72 Mean : 81.65 Mean :1.683 Mean :0.02539
## 3rd Qu.:32.10 3rd Qu.: 92.40 3rd Qu.:1.754 3rd Qu.:0.00000
## Max. :80.60 Max. :216.10 Max. :2.027 Max. :1.00000
## str dm ckd dental
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:1.0000
## Median :0.0000 Median :0.0000 Median :0.0000 Median :1.0000
## Mean :0.0088 Mean :0.1451 Mean :0.2517 Mean :0.8952
## 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000
## imp_id
## Length:83640
## Class :character
## Mode :character
##
##
##
```

```
write_csv(imp_nhanes, path = "~/Box/Analytics Team/Research/Research projects/SSB ban/imp_nhanes.csv")
```